IMM 383
July 1970

Courant Institute of
Mathematical Sciences

A Further Note on Top-down Deterministic Languages

Derick Wood

Prepared under Grant No. NSF-GJ-95
with the National Science Foundation

New York University

IMM 383
July 1970

New York University

Courant Institute of Mathematical Sciences

A FURTHER NOTE ON TOP-DOWN DETERMINISTIC LANGUAGES

Derick Wood

Results obtained at the Courant Institute of Mathematical Sciences, New York University, with the National Science Foundation, Grant No. NSF-GJ-95.

Abstract

Two new families of languages, the F(k) and U(k) languages, are introduced each of which is, in some sense, a generalization of top-down deterministic languages. This leads us to new characterizations of s-grammars and LL(1) languages. We include a characterization of the unambiguous context-free languages, generalizations of the equivalence relation on s-grammars to LL(k) grammars, and a summary of the non-closure results for LL(k), F(k) and U(k) languages, and it is shown that non-degenerate hierarchies exist for the families of F(k) and U(k) languages.

Introduction

A survey of the approaches to top-down deterministic languages has been given in Wood (1969a), however since that time some earlier unpublished work by Schorre (1965) and Tixier (1967) has come to light. The aim of this paper is to investigate their approach, relating it to the LL(k) languages of Lewis and Stearns (1968). At the same time we take this opportunity to generalize some results of Korenjak and Hopcroft (1966) and to include a survey of the non-closure results for LL(k) languages, some of which are new.

In Section 2 we introduce separability, in Section 3 f-separability and f-quasi-separability, in Section 4 we deal with generalizations of the equivalence relation for s-grammars and in Section 5 are found the various non-closure counter-examples.

2. Notation

We use $\emptyset$ to denote the empty set. A grammar G is a 4-tuple

$$G = (N, T, S, P)$$

where

N	is a finite set of nonterminal symbols,
T	is a finite set of terminal symbols
S in N	is the sentence symbol, and
P	is a finite set of rules (or productions) of the form $X \rightarrow x$, X in N and x in $(N \cup T)^*$. Let $V = N \cup T$. If $X \rightarrow x$ in P then x is an alternative of X.

In the usual manner the free monoid generated by a set of symbols A, is denoted by A^*, similarly $A^+ = AA^*$. ε denotes the empty word. We have the binary relations $\underset{G}{=>}$, $\underset{G}{\overset{+}{=>}}$, $\underset{G}{\overset{*}{=>}}$ or more usually $=>$, $\overset{+}{=>}$, $\overset{*}{=>}$, on words over V^*, which define derivations over G. The language generated by a word w is the set $\{x : w \overset{*}{=>} x,\ x \text{ in } T^*\}$, written both as $L(w)$ or $\tilde{w}$, in this paper. The language generated by the grammar G, denoted $L(G)$, is $L(S)$. The length of a word x in V^* is denoted by $|x|$, and is the number of symbols in x, $|\varepsilon| = 0$. For $k > 0$, for all x in V^* let $k{:}x$ be x if $|x| \leq k$, otherwise x_1 where $|x_1| = k$ and $x = x_1 x_2$.

We say a grammar is admissible if for all X in V there exist derivations $S \overset{*}{=>} uXv$ and $X \overset{*}{=>} x$, where u, v in V^*, x in T^*. Henceforth grammar means admissible grammar.

Definition

For $k > 0$, X in N is said to be LL(k) if for all u, v, v', u_1, u_1', x, x', such that

$$S \overset{*}{\Rightarrow} uXv \Rightarrow uxv \overset{*}{\Rightarrow} uu_1,$$
$$S \overset{*}{\Rightarrow} uXv' \Rightarrow ux'v \overset{*}{\Rightarrow} uu_1', \text{ and}$$
$$k{:}u_1 = k{:}u_1' \text{ then } x = x', \text{ where}$$

u, u_1, u_1' in T^*, v, v' in V^* and $X \to x$, $X \to x'$ in P. Similarly a grammar G is LL(k) if for all X in N X is LL(k). L is an LL(k) language if it is generated by some LL(k) grammar. Let $\mathcal{L}(k)$ be the family of LL(k) languages and $\mathcal{L} = \bigcup_{\text{all } k} \mathcal{L}(k)$.

Definition.

A grammar G is an ε-free grammar if for all $X \to x$ in P, x in V^+. An ε-free LL(1) grammar is an s-grammar, similarly an s-language is a language that can be generated by an s-grammar.

A set X is ε-free if ε not in X.

Definition.

A nonterminal X has a cycle (or is cyclic) if there exists a derivation $X \overset{+}{\Rightarrow} uXv$, u,v in V^*, $uv \neq \varepsilon$. If $uv = \varepsilon$ then X has a loop. A grammar G has a cycle (loop) if at least one nonterminal in G has a cycle (loop). If X, Y in N have cycles and there exists no derivation $X \overset{+}{\Rightarrow} uYv$, u,v in V^*, then X, Y are said to have disjoint cycles.

. Separability and Context-Free Grammars

[illegible] ([illegible]) introduced the concept of separability which is further explored by [illegible] ([illegible]). We extend the notion of separability in order to apply it to arbitrary context-free grammars. By "set" we will mean "a set of words over some terminal alphabet T."

Definition

The *left quotient* (*right quotient*) of a set X by a set Y is $\{u : vu \text{ in } X,\ v \text{ in } Y\}$ ($\{u : uv \text{ in } X,\ v \text{ in } Y\}$) denoted by $X \backslash Y$ (X/Y).

Definition

The *proper left quotient* of a set X by a set Y is $\{u : vu \text{ in } X,\ v \text{ in } Y,\ u \neq \varepsilon\}$ denoted by $X.\backslash Y$. Similarly we define *proper right quotient* denoted by $X/.Y$.

Definition

Let (A,B) denote an ordered pair of sets. We say (X,Y) is *separable* if $X.\backslash X \cap Y/.Y = \emptyset$. Let

$$I(X,Y) \text{ denote the set } X.\backslash X \cap Y/.Y \ .$$

Remarks

1) It follows that if ε is in a set X then

$$X - \{\varepsilon\} \subseteq X.\backslash X \text{ and } X - \{\varepsilon\} \subseteq X/.X \ .$$

2) If x in T^+ then $\tilde{x}.\backslash \tilde{x} = \tilde{x}/.\tilde{x} = \emptyset$.

Definition

A nonterminal X is separable if for all $X \to x$ in P, where x is in VV^+ for all Y,Z in V such that $x = uYZv$, u,v in V^*, $(\tilde{Y},\tilde{Z})$ is separable. A grammar is separable if all X in N are separable.

Lemma 1

Given a grammar G such that for some $X \to uYZv$ in P (Y,Z) is separable then it follows that:

(i) for all $Y \overset{+}{\Rightarrow} yW$, y in V^*, W in V, $(\tilde{W},\tilde{Z})$ is separable,

(ii) for all $Z \overset{+}{\Rightarrow} Uz$, z in V^*, U in V, $(\tilde{Y},\tilde{U})$ is separable,

(iii) for all W and for all U as above, $(\tilde{W},\tilde{U})$ is separable.

Proof:

We will prove (i) in detail, (ii) and (iii) follow similarly.

(i) Assume the contrary, then $I(\tilde{W},\tilde{Z}) \neq \emptyset$ therefore there exists at least one p in $I(\tilde{W},\tilde{Z})$. Thus there exists p_1 in $\tilde{W}$ such that p_1p is in $\tilde{W}$, therefore qp_1 and qp_1p are in $\tilde{Y}$ for some q in $\tilde{y}$. Therefore p is in $I(\tilde{Y},\tilde{Z})$ giving a contradiction. The result follows. ||

Corollary 2.

If G is a separable grammar then for all X,Y in V such that there exists a derivation $Z \overset{+}{\Rightarrow} uXYv$, (X,Y) is separable.

The converse of Lemma 1 obviously does not hold, however we do have:

[illegible]

Given a grammar G, A → [illegible] in P and for all [illegible] in P, u,v,w in V^*, [illegible] in V, [illegible] is separable then [illegible] is separable.

Definition

We say an n-tuple $(X_1,\ldots,X_n)$ is R separable for $n \geq 2$, if for all i, $1 \leq i < n$, $(X_i, X_{i+1},\ldots,X_n)$ is separable.

Definition

A nonterminal X is R separable if for all $X \rightarrow X_1 \ldots X_n$ in P, $n \geq 2$, $(\tilde{X}_1,\ldots,\tilde{X}_n)$ is R separable. A grammar is R separable if all X in N are R separable.

We now give a lemma that demonstrates the behavior of R separability under substitution.

Lemma 4.

Given a grammar G such that for some $X \rightarrow X_1 \ldots X_n$ in P, $n \geq 2$, $(\tilde{X}_1,\ldots,\tilde{X}_n)$ is R separable then it follows that:

(i) for all i, $1 \leq i < n$, for all $X_i \overset{+}{\Rightarrow} yY$, y in V^*, Y in V, $(\tilde{Y},\tilde{X}_{i+1},\ldots,\tilde{X}_n)$ is R separable.

(ii) if for some i, $1 \leq i \leq n$, $X_i \rightarrow Y_1 \ldots Y_m$ is in P, $m \geq 2$ $(\tilde{Y}_1,\ldots,\tilde{Y}_m)$ is R separable; then $(\tilde{Y}_1,\ldots,\tilde{Y}_m,\tilde{X}_{i+1},\ldots,\tilde{X}_n)$ is R separable.

Proof:

(i) follows directly from Lemma 1.

(ii) Assume the contrary, then there exists a j, $1 \leq j \leq m$, such that $(\tilde{Y}_j,\tilde{Y}_{j+1},\ldots,\tilde{Y}_m\tilde{X}_{i+1},\ldots,\tilde{X}_n)$ is not separable. Let

$\tilde{y}$ denote $\tilde{Y}_{j+1} \ldots \tilde{Y}_m$ and $\tilde{x}$ denote $\tilde{X}_{i+1} \ldots \tilde{X}_n$. This implies there exists a q in $I(\tilde{Y}_j, \tilde{y}\tilde{x})$, a p in $\tilde{Y}_j$ and an r in $\tilde{y}\tilde{x}$ such that pq is in $\tilde{Y}_j$ and qr is in $\tilde{y}\tilde{x}$. Now $r = r_1' r_2'$ where r_1' in $\tilde{y}$, r_2' in $\tilde{x}$.

(a) $qr = qr_1 r_2$ where qr_1 in $\tilde{y}$, r_2 in $\tilde{x}$.

If $|r_2| \geq |r_2'|$ let $r_2 = r_{21} r_2'$ then r_{21} in $I(\tilde{X}_i, \tilde{x})$ as both pqr_1' in $\tilde{Y}_j \tilde{y}$ and pqr_1 in $\tilde{Y}_j \tilde{y}$. If $|r_2| < |r_2'|$ a similar argument holds.

(b) $qr = q_1 q_2 r$ where q_1 in $\tilde{y}$, $q_2 r$ in $\tilde{x}$. As r_2' in $\tilde{x}$ we have $q_2 r_1'$ in $\tilde{x}/.\tilde{x}$ and as p,pq in $\tilde{Y}_j$, q_1, r_1' in $\tilde{y}$ we have $q_2 r_1'$ in $\tilde{Y}_j \tilde{y} . \backslash \tilde{Y}_j \tilde{y}$.

Both cases lead to $I(\tilde{X}_i, \tilde{x}) \neq \emptyset$, a contradiction. ||

Corollary 5.

If G is an R separable grammar then G is separable. This is really a corollary to the definition of R separability.

Corollary 6.

If G is an R separable grammar then for all X_i in V such that there exists a derivation $Z \overset{+}{\Rightarrow} X_1 \ldots X_n$, $n \geq 2$, $(X_1, \ldots, X_n)$ is R separable.

Definition

We say an n-tuple $(X_1, \ldots, X_n)$, $n \geq 2$, is L separable if for all i, $1 < i \leq n$, $(X_1 \ldots X_{i-1}, X_i)$ is separable. We can extend the definition to nonterminals and grammars.

We now give two examples which illustrate separability.

Example 1.

Let $X \rightarrow X_1X_2X_3$ be a rule of some grammar, where $\tilde{X}_1 = \{a\}^* = \tilde{X}_2$ and $\tilde{X}_3 = \{a\}$.

Then (X_1,X_2) and (X_2,X_3) are separable but

(X_1,X_2X_3) and (X_1X_2,X_3) are not separable.

Thus (X_1,X_2,X_3) is not L or R separable (note that X is ambiguous).

Example 2.

Let X be as above, where $\tilde{X}_1 = \{a,\varepsilon\}$, $\tilde{X}_2 = \{b,\varepsilon\}$ and $\tilde{X}_3 = \{a\}$. Then X is both L and R separable and therefore separable.

This leads to the following lemma.

Lemma 7.

A grammar is L separable if and only if it is R separable.

Proof: This is not included as it is very similar to that used in Lemma 4. ||

Definition

A grammar is s(uper)-separable if it is R and hence L separable.

Corollary 8.

If G is an s-separable grammar then for all X_i in V such that there exists a derivation $Z \overset{+}{\Rightarrow} X_1 \ldots X_n$, $n \geq 2$, $(\tilde{X}_i,\ldots,\tilde{X}_j)$ is s-separable for all i,j, $1 \leq i < n$, $i < j$.

Remark.

s-separability formalizes an intuitive notion of unambiguity, i.e. if $(\tilde{X},\tilde{Y})$ is separable then we know that there is no subword v, such that there exists x in $\tilde{X}$, y in $\tilde{Y}$ such that xv in $\tilde{X}$ and vy in $\tilde{Y}$. Therefore it follows that each word x in $\tilde{X}\tilde{Y}$ can be uniquely partitioned into x_1x_2 with x_1 in $\tilde{X}$, x_2 in $\tilde{Y}$. Thus s-separability is a necessary condition for unambiguity, that it is not sufficient is shown by the next example.

Example 3.

Let $G = (\{S,X_1,X_2,Y_1,Y_2\},\{a,b\},S,\{S\to X_1Y_1|X_2Y_2,$

$X_1\to a,\ X_2\to a,\ Y_1\to b,\ Y_2\to b\})$.

G is trivially s-separable and ambiguous.

We are now in a position to prove the following necessary and sufficient condition for unambiguity.

Theorem 9.

A grammar G is unambiguous if and only if the following property is satisfied:

(i) for all X in N, for all $X \to x_1$, $X \to x_2$ in P

$\tilde{x}_1 \cap \tilde{x}_2 = \emptyset$, $x_1 \neq x_2$.

(ii) G is s-separable. This is Property A.

Proof: As sufficiency follows trivially consider necessity. Assume the property holds and G is ambiguous. There exists at least one word x in L(G) such that x has at least 2 distinct derivation sequences; let these be $\{v_i\}$ and $\{w_i\}$, where

$v_0 = w_0 = X$ and $v_p = w_q = x$. There exists j, $0 \leq j < \min(p,q)$ such that

$$v_i = w_i, \quad 0 \leq i \leq j \quad \text{and} \quad v_{j+1} \neq w_{j+1}.$$

Let $v_j = aXv$, $v_{j+1} = ax_1v$, $w_{j+1} = ax_2v$, a in T^*, X in N, v in V^*, where $X \rightarrow x_1$, $X \rightarrow x_2$ in P, $x_1 \neq x_2$.

Now either $x_1 \overset{*}{\Rightarrow} a_1$, $x_2 \overset{*}{\Rightarrow} a_1$, $v \overset{*}{\Rightarrow} a_2$, a_1, a_2 in T^*, where

$x = a\, a_1 a_2$. Thus $\tilde{x}_1 \cap \tilde{x}_2 \neq \emptyset$

or $\quad x_1 \overset{*}{\Rightarrow} a_1$, $x_2 \overset{*}{\Rightarrow} a_1'$, a_1, a_1' in T^*, $|a_1| < |a_1'|$ say,

and $\quad v \overset{*}{\Rightarrow} a_2$, $v \overset{*}{\Rightarrow} a_2'$, a_2, a_2' in T^* with $a_1 a_2 = a_1' a_2'$.

Thus $I(\tilde{X}, \tilde{v}) \neq \emptyset$, which by Corollary 8 contradicts the second condition. The theorem follows. ||

<u>Definition</u>

G is a <u>binary grammar</u> if for all $X \rightarrow x$ in P, $x = \varepsilon$ or $x = X_1 X_2$, X in N, X_1, X_2 in V.

Schorre (1965) proved the following result.

<u>Corollary 10</u>.

A binary grammar is unambiguous if and only if the following property is satisfied:

(i) for all X in N, for all $X \rightarrow x_1$, $X \rightarrow x_2$ in P, $x_1 \neq x_2$, $\tilde{x}_1 \cap \tilde{x}_2 = \emptyset$.

(ii) G is separable. This is <u>Property B</u>.

<u>Proof</u>: (X,Y) is s-separable if and only if (X,Y) is separable. ||

Remark.

The proof of Theorem 9 can be followed through by replacing s-separabiiity with R separability, and thus Lemma 7 is not necessary for this proof. Thus a direct proof of Corollary 10 would be much shorter than that originally given by Schorre (1965). Property B can be applied whenever at most two nonterminals appear together in the right sides of the rules in a grammar, so we obtain the following result.

Definition

A grammar is in ε-(k,2)-normal form, $k \geq 1$ if for all $X \rightarrow x$ in P, either x in T^*, $|x| \leq k$ or $x = ay$, a in T^*, $|a|=k$, y in $N \cup NN$.

It is known (Wood, 1970), that every grammar can be put in ε-(k-2) normal form.

Corollary 11.

An ε-(k,2) normal form grammar is unambiguous if and only if Property B. holds.

Definition

A grammar is an operator grammar if for all $X \rightarrow x$ in P, x in $T\,(NT)^*$ (Greibach, 1965).

Corollary 12.

An operator grammar is unambiguous if for all X in N, for all $X \rightarrow x_1$, $X \rightarrow x_2$ in P, $x_1 \neq x_2$, $\tilde{x}_1 \cap \tilde{x}_2 = \emptyset$.

[illegible]. f-separability and k:separability

Definition

(X,Y) is f-separable if

$$f(X.\backslash X) \cap f(Y/.Y) = \emptyset ,$$

where f is a mapping from T^+ into some set D.

Letting $f(u) = u$ for all u in T^+ then we have separability. We extend the definition to give f-s-separability in the obvious way. We are particularly interested in the following special mapping.

Definition

(X,Y) is k:separable if $k:(X.\backslash X) \cap k:(Y/.Y) = \emptyset$. We define k:s-separable in a similar way; let $I_k(X,Y)$ denote the set $k:(X.\backslash X) \cap k:(Y/.Y)$.

Definition

Given a nonnegative integer k, a grammar G is a U(k) grammar if the following property is satisfied:

(i) for all X in N, for all $X \to x_1$, $X \to x_2$ in P, $x_1 \neq x_2$, $k:\tilde{x}_1 \cap k:\tilde{x}_2 = \emptyset$.

(ii) G is k:s-separable.

Definition

(X,Y) is f-q(uasi)-separable if

$f(X.\backslash X) \cap f(Y) = \emptyset$ where f is a mapping from T^+ into some set D. Similarly define k:q-separable and k:q-s-separable.

Definition

Given a nonnegative integer k, a grammar G is an F(k) grammar if the following property holds:

(i) for all X in N, for all $X \to x_1$, $X \to x_2$ in P, $x_1 \neq x_2$ $k:\tilde{x}_1 \cap k:\tilde{x}_2 = \emptyset$.

(ii) G is k:q-s-separable.

F(k) grammars are a generalization of the RCF (regular context free) grammars of Tixier (1967) and the FCR (first character recognition) grammars of Schorre (1965).

Definition

A language is a U(k) or F(k) language if it is generated by a U(k) or F(k) grammar. Let $\mathcal{U}(k)$ and $\mathcal{F}(k)$ denote the families of U(k) and F(k) languages. Further let $\mathcal{U}$ = {L: there exists a k > 0, such that L is in $\mathcal{U}(k)$} and similarly define $\mathcal{F}$.

Let us look at some basic properties of these families. From the definitions and Theorem 9 we have:

Corollary 13.

For all L, L in $\mathcal{U}$, L is unambiguous.

However, because of the definition of q-separability we have:

Letting $X = \{c,\varepsilon\}$ and $Y = \{ca^ib^ic^j,\ a^ib^jc^j : i,j \geq 1\}$ we have $\{ca^ib^ic^j,\ ca^ib^jc^j : i,j \geq 1\} \subset XY$ and (X,Y) is 2: q-separable. Further as both X and Y can be generated by 1: q-s-separable rules we obtain the following result.

Theorem 15.

$\mathcal{F}$ contains inherently ambiguous languages. We now exhibit a language which is not in $\mathcal{F}$.

Theorem 16.

$L_1 = \{a^i, a^i b^i : i,j \geq 1\}$ is not in $\mathcal{F}$, and therefore $\mathcal{F}$ is a proper subset of the family of context-free languages.

Proof:

If G is such that $L(G) = L_1$ then G must have two disjoint cycles and therefore at least three nonterminals. Otherwise $L(G)$ would contain words of the form $a^i b^j$, $i > j$, thus we need at least one nonterminal for each cycle and one nonterminal that branches to either cycle, let this be X. Then X has at least two alternatives, one of which leads to words of the form a^i and the other to words of the form $a^i b^i$; let these be $X \to x_1$ and $X \to x_2$. Then for any $k > 0$, $k:\tilde{x}_1 \cap k:\tilde{x}_2 \neq \emptyset$. ||

Corollary 16.

L_1 is not in $\mathcal{U}$.

It turns out that $\mathcal{U}$ and $\mathcal{F}$ are incomparable, an unexpected result.

Theorem 17.

$L_2 = \{a^i b c^j : i \geq j \geq 1\}$ is in $\mathcal{U}$ but not in $\mathcal{F}$.

Proof: The rules $S \to AB$, $A \to aA|\varepsilon$, $B \to abc|b$ are 1:s-separable; therefore L_2 is in $\mathcal{U}(1)$.

There are three distinct ways in which L_2 can be generated,

these are:

(a) form $a^i a^j b\ c^j$

(b) form $a^j a^i b\ c^j$

(c) form $a^j (b|\varepsilon)^j$.

Each of these can be shown to be non-generable by an F(k) grammar for any $k > 0$; in a similar manner to the proof of Theorem 15. We will prove (a) only. There must be two disjoint cycles, one to generate a^i, one to generate $a^j b\ c^j$ and there must be a nonterminal that branches to both cycles; so that $a^i a^j b\ c^j$ can be formed. Let this be the rule $X \to AB$, where $\tilde{A} = \{a^i\}$, $\tilde{B} = \{a^i b\ c^i\}$ then $\tilde{A}.\backslash \tilde{A} = \{a^i\}$ and therefore

$$k{:}\ (\tilde{A}.\backslash \tilde{A}) \cap k{:}\ \tilde{B} \neq \emptyset \text{ for any } k > 0. \qquad ||$$

We now show every LL(k) grammar is U(k).

<u>Theorem 18</u>.

For all $k > 0$, each LL(k) grammar is U(k).

<u>Proof</u>:

Each LL(k) grammar trivially satisfies the first condition for a grammar to be U(k), therefore it remains to show that each LL(k) grammar is k:s-separable. Proceed by contradiction. Given G, an LL(k) grammar assume it is not a U(k) grammar. Then there exists at least one rule $Y \to xXy$ such that $(\tilde{X},\tilde{y})$ is not k:separable. This implies: b in $I_k(\tilde{X},\tilde{y})$.
If bc in $I(\tilde{X},\tilde{y})$, c in T^* then G is ambiguous, therefore we must have

[illegible] in [illegible] and [illegible] in [illegible], [illegible] in [illegible]*,

[illegible] $=$ [illegible] and [illegible] $\neq$ [illegible].

[illegible]

[illegible] in $\tilde{X}$ for some [illegible] in $\tilde{X}$ and

[illegible] in y for some [illegible] in $\tilde{y}$.

We have

$$Y \Rightarrow xXy \overset{*}{\Rightarrow} u_1Xy \overset{*}{\Rightarrow} u_1ay, \quad u_1 \text{ in } T^*, \text{ and} \qquad (1)$$

$$Y \Rightarrow xXy \overset{*}{\Rightarrow} u_1Xy \overset{*}{\Rightarrow} u_1abcy, \qquad (2)$$

which in turn implies there exists Z in N such that

$$X \overset{*}{\Rightarrow} a_1Z \Rightarrow a_1z_1 \overset{*}{\Rightarrow} a_1a_2 \qquad (3)$$

$$X \overset{*}{\Rightarrow} a_1Z \Rightarrow a_1z_2 \Rightarrow a_1a_2bc\ , \qquad (4)$$

where

$a_1a_2 = a$, a_1, a_2 in T^*, $z_1 \neq z_2$, $Z \to z_1$, $Z \to z_2$ in P.
Thus letting $S \overset{*}{\Rightarrow} uYv$, combining (1) with (3) and (2) with (4) and noticing that in (1) $y \overset{*}{\Rightarrow} bde$, we have that Z is not LL(k). ||

This leads to an indirect proof of:

<u>Corollary 19</u>.

Each LL(k) grammar is unambiguous.

By an almost identical proof, we also have:

<u>Corollary 20</u>.

For all $k > 0$, each LL(k) grammar is F(k).

We now compare $\mathcal{L}(1)$, $\mathcal{U}(1)$, $\mathcal{F}(1)$.

Theorem 21.

(i) $\mathcal{L}(1) = \mathcal{F}(1)$

(ii) $\mathcal{L}(1) \neq \mathcal{U}(1)$.

Proof:

(i) Because of Corollary 20 we need only show that 1:q-s-separability implies LL(1). Assume it does not. Then given G, an F(1) grammar there exists at least one X in N which is not LL(1). Thus

$$S \overset{*}{\Rightarrow} uXv \Rightarrow ux_1v \overset{*}{\Rightarrow} uu_1 ,$$

$$S \overset{*}{\Rightarrow} uXv' \Rightarrow ux_2v' \overset{*}{\Rightarrow} u\, u_1' , \quad u, u_1, u_1' \text{ in } T^*,$$

$$v,v' \text{ in } V^*, \quad X \rightarrow x_1, \quad X \rightarrow x_2 \text{ in } P,$$

$$1{:}\ u_1 = 1{:}\ u_1' \text{ but } x_1 \neq x_2 .$$

(a) ε in $\tilde{X}$.

Then $u_1 = \varepsilon$ and $u_2 = \varepsilon$ implies ε in $\tilde{x}_1 \cap \tilde{x}_2$ and $\tilde{X} - \{\varepsilon\} \subseteq \tilde{X}.\backslash\tilde{X}$ therefore $1{:}u_1$ in $1{:}(\tilde{X}.\backslash\tilde{X})$ and $1{:}u_1'$ in $1{:}\tilde{v}$

(b) ε not in $\tilde{X}$.

$1{:}\ \tilde{x}_1$ or $1{:}\ \tilde{x}_2 \neq \emptyset$.

In both cases we have a contradiction.

(ii) $L_3 = \{a^i(b\ bbd)^i{:}\ i \geq 1\}$ is known to be LL(2) but not LL(1) (Rosenkrantz and Stearns, 1970). It can be generated by a U(1) grammar, however. Let

$$G = (\{S,A,B,C\},\{a,b,d\},S,P)$$

where

$P = \{S \to aA,$

$\qquad A \to aAB|B,$

$\qquad B \to bC,$

$\qquad C \to bd|\varepsilon\}.$

As $\tilde{E}/.\tilde{E} = \emptyset$ the grammar is trivially U(1), but not F(1). ||

We have the weaker characterization:

<u>Corollary 22</u>.

A grammar is an s-grammar iff it is an ε-free U(1) grammar.

Because L_1 in Theorem 15 is deterministic but not in U we are lead to the following result.

<u>Theorem 23</u>.

U and the family of deterministic languages are incomparable.

<u>Proof</u>:

Let $L_3 = \{ww^R: w \text{ in } \{a,b\}^*\}$, then L_3 is not deterministic but it is generated by the rules

$$S \to aSa|bSb|\varepsilon \ ,$$

which are trivially U(1) and F(2). ||

<u>Corollary 24</u>.

$$\mathcal{F} \cap \mathcal{U} \supset \mathcal{L} \ .$$

We now show that a nontrivial hierarchy exists for both the $\mathcal{F}(k)$ and $\mathcal{U}(k)$ families.

<u>Theorem 25</u>.

For any $k \geq 1$, $L_4(k) = \{a^i(b|b^{k+1}d)^i: i \geq 1\}$ is not F(k).

Proof:

Any grammar generating $L_4(k)$ must have a cycle for some nonterminal X, say, i.e.

$$X \stackrel{+}{=>} uXv$$

where $u \stackrel{*}{=>} a^m$, $v \stackrel{*}{=>} (b|b^{k+1}d)^m$, $m \geq 1$. This implies: $b^k d$ in $\tilde{X}.\backslash\tilde{X}$ and b^k in $k:\tilde{v}$. Thus

$$k: (\tilde{X}.\backslash\tilde{X}) \cap k:\tilde{v} \neq \emptyset.$$

However it is $F(k+1)$. ||

Corollary 26.

$$\mathcal{F}(i) \subset \mathcal{F}(i+1) \text{ for all } i \geq 1.$$

Theorem 27.

For any $k \geq 1$, $L_5(k) = \{a^i(d|b^k d|db^{k+1})^i : i \geq 1\}$ is not $U(k)$.

Proof:

As in Theorem 25 we find an X in N such that $X \stackrel{+}{=>} uXv$, b^{k+1} in $\tilde{X}.\backslash\tilde{X}$ and b^k in $\tilde{v}/.\tilde{v}$; thus $k:(\tilde{X}.\backslash\tilde{X}) \cap k:(\tilde{v}/.\tilde{v}) \neq \emptyset$; therefore $L_5(k)$ is not $U(k)$, but it is $U(k+1)$. ||

Operations on U(k) and F(k) languages.

By noting that $L_1 = \{a^i, a^i b^i : i \geq 1\}$ is neither $U(k)$ nor $F(k)$ for any $k > 0$, but that it is the union of two languages which are both $U(1)$ and $F(1)$, we can construct the various non-closure results given in Table 1, in a similar manner to those given for LL(k) languages in Section 5.

4. Relations and Context-Free Grammars

Korenjak and Hopcroft (1966) (henceforth let KH denote this reference) solved the equivalence problem for s-grammars by considering a relation on V^+ (in fact, an equivalence relation). Attempts since then have been made, without success, to extend this notion to solve the equivalence problem for LL(k) grammars, although this problem has been solved by a different route in Rosenkrantz and Stearns (1970). Tixier extended the relation to LL(1) grammars; we now extend it to unambiguous grammars and languages.

Definition.

A nonterminal X in N has the prefix property if $X \overset{*}{\Rightarrow} x = yz$ $(z \neq \varepsilon)$ then $X \overset{+}{\not\Rightarrow} y$.

The following result is given in KH.

Lemma 28.

Given G, an s-grammar, every X in N has the prefix property.

If s has the prefix property then $L(G)$ is said to have the prefix property. Therefore every s-language has the prefix property.

Definition.

If X is a set of words let $sh(X)$ denote the length of a shortest word in X.

Definition.

A grammar G is in (1,2)-normal form if for all $X \to x$ in P either x in T, x in TN or x in TNN.

Definition.

If X, Y are sets of words from T^* and fall all x, x in X if and only if x in Y then write $X \equiv Y$ (i.e. $X \equiv Y$ if and only if $X = Y$). We extend this to the catenation product of sets of words and to words over V (i.e. $x \equiv y$ if and only if $\tilde{x} \equiv \tilde{y}$). We say $X \equiv Y$ is an equivalence pair.

As in KH we have:

Lemma 29.

The relation '$\equiv$' is a congruence relation under catenation product.

Proof:

(i) reflexive: $X \equiv X$ as $X = X$

(ii) symmetric: $X \equiv Y$ implies $Y \equiv X$

(iii) transitive: $X \equiv Y$ and $Y \equiv Z$ implies $X \equiv Z$

(iv) catenation: $X \equiv Y$ and $W \equiv Z$ implies $XW \equiv YZ$.

Considering the corresponding sets of words (i)-(iv) follow trivially. ||

Remark

Note that the above lemma holds for any sets of words, however they are generated.

We now assume that for any set X, X is nonempty.

Definition

We say $\Pi(X_1,\ldots,X_n)$ iff $(X_1,\ldots,X_n)$ is s-separable.

Lemma 30.

If $\Pi(X,Z)$, $\Pi(W,Y)$, $W \equiv X$ and $WY \equiv XZ$ then $Y \equiv Z$.

If:

Assume otherwise. Then there exists a shortest word a that contradicts $Y \equiv Z$, without loss of generality assume a in Y.

Let t be a shortest word in W (and hence in X); then

ta in WY and ta in XZ.

Now no proper prefix of t can be in X, by construction, therefore ta_1 in X, a_2 in Z, $a_1a_2 = a$, $a_1 \neq \varepsilon$. Further a_2 in Y as a is a shortest word that contradicts $Y \equiv Z$. Therefore t, ta_1 in W, a_1, a_2, a_2 in Y contradicts $\Pi(W,Y)$. ||

__Corollary 31__. __Left cancellation__.

If $\Pi(W,Z)$, $\Pi(W,Y)$ and $WY \equiv WZ$ then $Y \equiv Z$.

__Corollary 32__.

If $\Pi(X,Z)$, $\Pi(W,Y)$, $Y \equiv Z$ and $WY \equiv XZ$ then $W \equiv X$.

__Corollary 33__. __Right cancellation__.

If $\Pi(W,Y)$, $\Pi(X,Y)$ and $WY \equiv XY$ then $W \equiv X$.

A useful operation is substitution of equivalences, which preserves equivalences.

__Lemma 34__.

If $\Pi(X,Y)$, $\Pi(W,Y)$ and $XY \equiv Z$ then $X \equiv W$ iff $WY \equiv Z$.

__Proof__:

__If__: $XY \equiv Z$ and $WY \equiv Z$ then $XY \equiv WY$ and $X \equiv W$.

__only if__: $X \equiv W$ and $Y \equiv Y$ then $XY \equiv WY$ giving $WY \equiv Z$. ||

KH give two transformations on equivalence pairs; we extend these in a natural way.

Definition. The A-transformation.

Given an ε-free set X we let X(a) denote the subset of X defined as $\{u_1 : u \text{ in } X,\ u = au_1\}$.

Given the equivalence pair $X_1 \ldots X_n \equiv Y_1 \ldots Y_m$ form $X_1(a)$ and $Y_1(a)$ for all a in T. We replace the equivalence pair by a set of new equivalence pairs

$X_1(a)X_2 \ldots X_n \equiv Y_1(a)Y_2 \ldots Y_m$ for all a in T.

We have trivially:

Lemma 35.

In the above definition

$$X_1 \ldots X_n \equiv Y_1 \ldots Y_m \text{ iff for all a in T}$$
$$X_1(a)\ X_2 \ldots\ X_n \equiv Y_1(a)\ Y_2 \ldots\ Y_m.$$

Remarks.

(i) Note that $X(a) = X \backslash \{a\}$.

(ii) If the equivalence pair is over $N^+ \times N^+$ then the A-transformation is carried out as a left substitution, followed by a collecting of terms, as more than one alternative of X_1 may begin with a specific terminal symbol.

(iii) Note that $X_1(a)$ or $Y_1(a)$ for some a in T may contain the empty word. For s-languages (and grammars) we have ε in $X_1(a)$ iff ε in $Y_1(a)$ for any a in T. However this is obviously not true for LL(k) languages (and grammars) in general.

We now generalize the B-transformation of KH.

<u>Definition</u>. <u>The B-transformation</u>.

Let $X_1 \ldots X_n \equiv Y_1 \ldots Y_m$, a in X_1 and $aZ \subseteq Y_1 \ldots Y_\ell$, $\ell \geq 1$, for some set Z such that there exists no set Z_1, $Z \subset Z_1$ and $aZ_1 \subseteq Y_1 \ldots Y_\ell$, then replacing

$$X_1 \ldots X_n \equiv Y_1 \ldots Y_m \quad \text{by}$$

$$X_2 \ldots X_n \equiv ZY_{\ell+1} \ldots Y_m \quad \text{and} \quad X_1 Z \equiv Y_1 \ldots Y_\ell$$

we have <u>the B-transformation</u>.

<u>Theorem 36</u>.

$$X_1 \ldots X_n \equiv Y_1 \ldots Y_m \quad \text{iff}$$

$$X_2 \ldots X_n \equiv ZY_{\ell+1} \ldots Y_m \quad \text{and} \quad X_1 Z \equiv Y_1 \ldots Y_\ell .$$

<u>Proof</u>: As in KH.

Further, s-separability is preserved. $||$

Considering equivalences on s-grammars we have the following corollaries.

Noting that Z will have the form $Z_1 \ldots Z_p$, Z_i in N, $p \geq 0$, we have:

<u>Corollary 37</u>.

$0 \leq |p| \leq |a| + 1$, if $Y_1 \ldots Y_q \not\overset{+}{\Rightarrow} ay$, for $q < \ell$.

<u>Definition</u>.

Given a grammar G_i, let

$t_i = \max(\{sh(X) : X \text{ in } N_i\})$ and let $t = \max(\{t_i\})$, all i.

This leads to the following corollary.

Corollary 38.

If $X_1 \ldots X_n \equiv Y_1 \ldots Y_m$ is an equivalence pair on s-grammars G_1 and G_2, then

(i) $sh(X_1 \ldots X_n) = sh(Y_1 \ldots Y_m)$

(ii) $1 \leq m \leq nt$, i.e. the length of the right side is bounded by the length of the left side.

Then we have the special case of KH.

Corollary 39.

If $n \leq t+3$ then the left sides generated by the B-transformation have length at most $t+2$, and therefore the right sides have length at most $t(t+2)$.

Definition.

Given two sets X,Y then we say:

(i) $X \cdot\!< Y$, X is left string contained in Y, if for all x in X, xy in Y for some y in T^*,

(ii) $X \cdot\!> Y$, X left string contains Y, if for all x in X, there exists y in Y and z in T^* such that $x = yz$,

(iii) $X <\!\cdot\, Y$, X is right string contained in Y, if for all x in X, yx in Y for some y in T^*,

(iv) $X >\!\cdot\, Y$, X right string contains Y, if for all x in X, there exists y in Y and z in T^*, such that $x = zy$.

Similarly we can define $x \cdot\!<, \cdot\!>, <\!\cdot, >\!\cdot\; y$ for words x,y in T^*.

We now have the following lemma.

Lemma [illegible].

Given $X_1X_2 = Y_1Y_2$, where [illegible] the prefix property, we have:

(i) If $sh(X_1) = sh(Y_1)$ then $X_1 \equiv Y_1$.

(ii) If $sh(X_1) < sh(Y_1)$ then $X_1 \cdot< Y_1$.

(iii) If $sh(X_1) > sh(Y_1)$, then $X_1 \cdot> Y_1$.

Proof:

Let $\bar{x}_1, \bar{x}_2, \bar{y}_1, \bar{y}_2$ be shortest words in X_1, X_2, Y_1, Y_2, and $\bar{X}_1, \bar{X}_2, \bar{Y}_1, \bar{Y}_2$ be the corresponding sets of shortest words.

(i) Choose the shortest word which contradicts $X_1 \equiv Y_1$, let this be x_1. It follows that

$$\bar{X}_1 = \bar{Y}_1 \quad \text{and} \quad \bar{X}_2 = \bar{Y}_2.$$

Now $x_1\bar{x}_2$ in Y_1Y_2. Thus we have an immediate contradiction, therefore $X_1 \equiv Y_1$.

(ii) We have $\bar{X}_1 \cdot< \bar{Y}_1$ and $\bar{X}_2 >\cdot \bar{Y}_2$.

Let x_1 be the shortest word in X_1 such that there exists no y in T^* such that x_1y in Y_1. Immediately we have

$x_1\bar{x}_2$ in Y_1Y_2 implies x_{11} in Y_1, $x_{12}\bar{x}_2$ in Y_2

where $x_{11}x_{12} = x_1$, giving

$x_{11}\bar{y}_2$ in X_1X_2 and as $\bar{X}_2 >\cdot \bar{Y}_2$ we have

v_1 in X_1, $v_2\bar{y}_2$ in X_2, where $v_1v_2 = x_{11}$, giving

a contradiction of the prefix condition.

(iii) Is proved in a similar way to (ii). ||

We have in fact a stronger result.

Theorem 41.

Given sets X_1, X_2, Y_1, Y_2 with the prefix property, $sh(X_1) = sh(Y_1)$ and $sh(X_2) = sh(Y_2)$ then

$$X_1X_2 \equiv Y_1Y_2 \quad \text{iff} \quad X_1 \equiv Y_1 \quad \text{and} \quad X_2 \equiv Y_2.$$

The following question arises.

Does Theorem 41 hold without the prefix property?

Example 1.

Let $X_1 = \{ab^i : i \geq 0\}$, $X_2 = \{d\}$

$Y_1 = \{a\}$, $Y_2 = \{b^i d : i \geq 0\}$

X_1 does not have the prefix property.

$sh(X_1) = sh(Y_1)$, $sh(X_2) = sh(Y_2)$ and $X_1X_2 \equiv Y_1Y_2$

but $X_1 \not\equiv X_2$ and $Y_1 \not\equiv Y_2$.

The most we can say is:

Lemma 42.

Given sets $X_1, X_2, Y_1\ Y_2$ with $sh(X_1) = sh(Y_1)$ and $sh(X_2) = sh(Y_2)$ then $X_1X_2 \equiv Y_1Y_2$ only if $X_1 \equiv Y_1$ and $Y_1 \equiv Y_2$.

This is just a trivial restatement of the catenation property.

[illegible]

Comparison of [illegible], [illegible], [illegible] and Deterministic Languages

Operation	IL(x)	[illegible]	[illegible]
Union	no	no	no
Concatenation	no	no	no
Concatenation with R	no	yes	no
Closure	no	no	no
Reversal	no	no	no
Intersection	no	no	no
Complement	no	yes	[illegible]
Intersection with regular set	no	yes	no
Substitution	no	no	no
ε-free substitution	no	no	no
Gsm mappings	no	yes	no
ε-free gsm mappings	no	yes	no
Inverse deterministic gsm mappings	no	yes	no
Quotient with regular set	no	yes	[illegible]
Homomorphism	no	no	no
ε-free homomorphism	no	no	no

5. Operations and LL(k) Languages

In Table I we compare LL(k) languages with the deterministic context-free languages of Ginsburg and Greibach (1966). The results for LL(k) languages are non-closure results; this with the known result (Rosenkrantz and Stearns, 1970) that the LL(k) languages form the largest known class for which the equivalence problem is decidable, their misbehavior is surprising. Most of the results detailed below appeared previously in Korenjak and Hopcroft (1966), Rosenkrantz and Stearns (1970), Tixier (1967) and Wood (1969b).

Boolean Operations.

Lemma 43.

$\mathcal{L}$ is not closed under (i) union, (ii) intersection, (iii) complement.

Proof:

(i) Let $L_1 = \{a^i : i \geq 1\}$, $L_2 = \{a^i b^i : i \geq 1\}$
then $L_1 \cup L_2$ is not LL(k) for any $k > 0$.

(ii) Let $L_3 = \{a^i (b|c)\ a^i (b|c) : i \geq 1\}$,
$L_4 = \{a^i b\ a^j b,\ a^i c\ a^j c : i,j \geq 1\}$,
then $L_5 = L_3 \cap L_4 = \{a^i b\ a^i b,\ a^i c\ a^i c : i,j \geq 1\}$
which is not LL(k) for any k.

(iii) Let $L_6 = \{a^i b^j : j \geq i \geq 1\}$
then $\{a,b\}^* - L_6$ is not LL(k) for any k,
this is proved in Rosenkrantz and Stearns (1970). ||

Because L_1 and L_4 are regular sets we have the following.

Corollary 44.

$\mathcal{L}$ is not closed under union or intersection with a regular set.

Letting $L_7 = \{a^i b a^j c, a^i c a^j b : i,j \geq 1\}$ then

$L_3 - L_7 = L_6$ and as $\{a,b\}^* - L_6$ is not LL(k) we have:

Corollary 45.

$\mathcal{L}$ is not closed under subtraction or subtraction with a regular set.

Mappings.

Let $L_8 = cL_1 \cup dL_2$, then L_8 is LL(1).
Define a homomorphism σ, that maps d onto c and the other symbols onto themselves, then

$\sigma(L_8) = L_9 = \{ca^i, ca^i b^i : i \geq 1\}$ which is not LL(k) for any k. We have shown

Lemma 46.

$\mathcal{L}$ is not closed under ε-free homomorphism.
Further as homomorphism is a special case of substitution we have

Corollary 47.

$\mathcal{L}$ is not closed under ε-free(finite) substitution, homomorphism or (finite) substitution.
Similarly we can define a gsm mapping that performs the homomorphism σ, therefore we also have

Corollary 48.

$\mathcal{L}$ is not closed under ε-free gsm mappings.

Even when these non-closure results hold it is usual for inverse mappings to give closure, however we have:

Lemma 49.

$\mathcal{L}$ is not closed under inverse ε-free homomorphism or inverse ε-free gsm mappings.

Proof:

Define the homomorphism σ_1 (gsm mapping) that maps a and b onto themselves and c onto b.

Then $\sigma_1(L_5) = \{a^i b\ a^i b : i \geq 1\}$ which is LL(k) but L_5 is not LL(k). ||

Finally we note that as $\mathcal{L}$ is not closed under ε-free homomorphism, it is not closed under k-limited erasing.

Products and Quotients.

Let $L_{10} = \{c\} \cup cL_2$ which is LL(1); then

$$L_{10}L_1 = \{ca^i,\ ca^i b^i a^j : i,j \geq 1\} \text{ is not LL(k).}$$

Trivially letting $L_{11} = L_1 \cup cL_2$ we have

$$\{c,cc\}L_{11} = \{ca^i, cca^i,\ cca^i b^i,\ ccca^i b^i : i \geq 1\} \text{ is not LL(k).}$$

We have just shown the following.

Lemma 50.

$\mathcal{L}$ is not closed under product, pre-product with a regular set or post-product with a regular set.

We have in fact shown a stronger result for RL, namely:

Corollary [illegible].

$\mathcal{L}$ is not closed under pre-product with a finite set.

However we do have (the proof left to the reader)

Lemma [illegible].

$\mathcal{L}$ is closed under post-product with a finite set.

Let $L_{12} = L_1 \cup cL_2 \cup \{c\}$ we have that

L_{12}^* contains strings of the form

$ca^n x$ and $ca^n c a^n y$ which means that L_{12}^* is not LL(k);

this gives:

Lemma 53.

$\mathcal{L}$ is not closed under catenation closure.

$L_{11} \backslash \{c, \varepsilon\} = L_{11} \cup L_2$ and is therefore not LL(k) giving

an expected result:

Lemma 54.

$\mathcal{L}$ is not closed under left quotient, left quotient with a regular set or left quotient with a finite set.

Tixier (1967) has shown the strongest possible result:

Lemma 55.

$\mathcal{L}$ is closed under left quotient with a single word.

Letting $L_{13} = L_2 / (\{ab^i : i \geq 1\} \cup \{\varepsilon\})$ this gives

$\{a^i b^i, a^{i+1} : i \geq 1\}$ which is not LL(k) for any k.

Therefore we have

Lemma 56.

$\mathcal{L}$ is not closed under right quotient or right quotient with a regular set.

However, as expected by the postproduct result:

Lemma 57.

$\mathcal{L}$ is closed under right quotient with a finite set.

We now examine those operations introduced by Ginsburg and Greibach (1965) which preserve the deterministic languages.

Definition.

$\mathrm{Init}(L) = \{u : uv \in L \text{ for some } v \in T^*\}$, thus

$\mathrm{Init}(L) = L/T^*$, the set of all initial subwords of words in L.

Taking $L = \{a^i b^i : i \geq 1\}$, $\mathrm{Init}(L) = \{a^i b^j : i \geq j \geq 0\}$

which is not LL(k) for any k, giving

Lemma 58.

Init does not preserve the LL(k) condition.

Definition.

$\mathrm{RDIV}(L_1, L_2) = \{u : uL_2 \subseteq L_1\}$.

Lemma 59.

RDIV does not preserve $\mathcal{L}$.

Proof:

Let $L_1 = \{a^{i+1} b^i (ab^j c | c) : i,j \geq 1\}$ and

$L_2 = \{a\, b^i c : i \geq 1\}$

then $\mathrm{RDIV}(L_1, L_2) = \{a^i, a^{i+1} b^i : i \geq 1\}$ which is not in $\mathcal{L}$.||

Miscellaneous Operations.

Lemma 60.

$\mathcal{L}$ is not closed under reversal.

Proof: Let $L = (L_1 \cup L_2)^R$ which is LL(1) but L^R is not LL(k).||

References

Ginsburg, S., and Greibach, S. A. (1966).
Deterministic context-free languages.
Information and Control 9, No. 6, pp. 620-648.

Greibach, S. A. (1965).
A new normal form theorem for context-free phrase structure grammars. JACM 12, No. 1, pp. 42-52.

Korenjak, A., and Hopcroft, J. E. (1966).
Simple determinstic languages, in Seventh Annual IEEE Switching and Automata Theory Conference, pp. 36-46.

Lewis II, P. M., and Stearns, R. E. (1968).
Syntax-directed transduction. JACM 15, No. 3, pp. 464-488.

Rosenkrantz, D. J. (1970).
Private communication.

Rosenkrantz, D. J., and Stearns, R. E. (1970).
Properties of deterministic top-down grammars, Information and Control, to appear.

Schorre, D. V. (1965).
A necessary and sufficient condition for a context-free grammar to be unambiguous. SDC Document SP-2153.

Tixier, V. (1967).
Recursive functions of regular expressions in language analysis. Doctoral dissertation, Computer Science Dept., Stanford University, CS 58.

Wood, D. (1969a).
A note on top-down deterministic languages.
BIT 9, No. 4, pp. 387-399.

Wood, D. (1969b).

The theory of left factored languages: Parts I & II. Computer Journal 12, No. 4, pp. 349-356 and 13, No. 1, pp. 55-62.

Wood, D. (1970).

A generalized normal form theorem for context-free grammars, Computer Journal, to appear.